論　説

Necessity of a Standardised Language Proficiency Test to Measure Participants' Proficiency and Compare Research Results in EFL Listening*

——リスニングにおける実験協力者の能力測定および研究結果の比較を目的とした標準テストの必要性——

Marisa Ueda†

Abstract

This article explores the importance and the necessity of standardised language proficiency tests in the field of EFL listening not only both prior to and after any experimental research but also to categorise and define participants proficiency level. Currently, a standardised language proficiency test is not widely used in many experimental researches in the field of EFL listening. Participants' level of competence in EFL listening is often subjectively judged by an instructor based on a faculty that the partcipnats belong to, the number of years that they had studied English and so on. Thus, the participants categorised as intermediate in one experimental research maybe categorised as advanced in another experimental research. Even a standardised language proficiency test is used to categorise and define participants proficiency level, it is not always clearly described *how* the levels of the participants are categorised and defined.

With a standardised language proficiency test, participants' proficiency level in EFL listening is objectively possible to measure prior to any experimental research. It also enables to compare many different research results objectively. With the results compared objectively, it enhances generalisability of the findings in the field of EFL listening.

Key words: listening, EFL (English as a foreign language), standardised language proficiency tests

---

*This article is based on Chapter 2 in a Ph. D. thesis submitted to Osaka University in November, 2014.

†Professor, Ph. D. in Language and Culture, Faculty of Economics, Ritsumeikan University, Japan. marisa@ec.ritsumei.ac.jp

This article explores the importance of standardised language proficiency tests, which are still the subject of wide-ranging debate amongst researchers. In this article, standardised language proficiency tests are defined as any language proficiency tests in English which are reliable, international, popular, relatively easy to access, capable of being compared and converted to other tests. Such tests are crucial for two reasons: To categorise and define participants' proficiency levels *before* an experimental research, and to evaluate an effect of the experiment objectively and scientifically test before and after the experimental research. The majority of previous studies on EFL/ESL listening strategies have compared more- and less-successful listeners. As Wu (1998) indicates, the preferred method in the field of testing assessment is to compare the top and bottom groups, ranging from 25% to 33% of the samples since the gap between the two groups is prominent and the differences are easy to compare. Therefore, it is quite natural to compare these two groups.

However, these studies contain significant variations and ambiguities, and employ only a limited number of standardised language proficiency tests for classification purposes before the study starts. Few standardised language proficiency tests are conducted for the purpose to determine the proficiency level of the participants *before* experiments. Researchers' classifications of learners as more- or less-proficient listeners vary significantly across studies and the lack of a standardised measure of listening proficiency can also diminish the overall generalisability of the findings since it cannot ensure that each study measures the same or common parameters. The studies which do not utilise a standardised language proficiency test to measure participants' proficiency and which cannot compare the results with the results of other researches can be divided into the following four types:

1) a study without any language proficiency tests,
2) a study with a language proficiency test which does not measure listening comprehension but other skills such as reading and mathematical skills,
3) a study with a local/minor language proficiency test and
4) a study with a standardised language proficiency test but without description/definition regarding a participants' proficiency classification.

Without the use of a standardised language proficiency test, those categorised as more-successful listeners in one study might be considered as intermediate in another, whilst those categorised as intermediate in one study might be classified as less successful in another. For example, the study of O'Malley et al.'s study (1989) is one of the first experimental studies on language learning strategies instruction which compare more- and less-effective listeners and is also described as a milestone study on listening strategy. They revealed that the mental processes of the students in listening comprehension actually par-

allel Anderson's (2010)[1] cognitive psychology theory in the following four ways:

1) The students are listening for larger chunks, shifting their attention to individual words only when there is a breakdown in comprehension;
2) They utilise both top-down and bottom-up processing strategies, whereas ineffective listeners repeatedly attempt to determine the meanings of individual words;
3) They are adept at constructing meaningful sentences from the input received, even though the meaning slightly differ from that of the actual text and
4) They apply their knowledge in three areas. (i.e. world knowledge, personal knowledge and self-questioning)

However, no standardised language proficiency test is used to define their competence in English *prior* to the experiment. The participants' proficiency levels are defined by a local test which is actually a school district placement test. There are additional studies in which language proficiency tests have not been used to assess the proficiency level of the participants *before* the experiment and even if a test has been used, in most cases, it is very minor or local to provide objective information about participants' comprehension levels. For example, Vandergrift (1997) reports the following three things:

1) clear differences in the listening strategies of 21 French listeners (i.e. 10 successful and 11 unsuccessful) based on four variables (i.e. level of language proficiency, gender, listening ability and learning style),
2) the use of metacognitive strategies such as comprehension monitoring, problem identification and selective attention seem to be the key factors that distinguish the successful listeners from the less successful ones and
3) the difference for gender is minimal.

The participants were chosen for the think-aloud sessions. Selection was based on both reported strategy use in Phase I and consultation with the teachers. Those reporting the greatest frequency, variety and sophistication of strategy use were classified as more successful listeners. Conversely, those at the other end of the continuum, reporting the least frequency, variety and sophistication of strategy use, were classified as less successful listeners. This selection and subsequent grouping were further corroborated by the participants' teachers on the basis of academic performance. Participants were interviewed individually in French using the ACTFL[2] Oral Proficiency Interview (Vandergrift, 1997, p. 390).

Although this is a report on listening comprehension strategies in a foreign language listening, the participants were merely measured with an *oral* proficiency interview. In addition, how they were categorised as successful and unsuccessful was stated but not objec-

tively clear so that there is no way for the other researchers to know the competence level of these participants in EFL listening itself.

As another example which no standardised language proficiency test is used, Graham, Santos and Vanderplank (2008) claim that strategy development seems to be related to proficiency levels. Their results show a high degree of stability of strategy use over six months, especially between the high and low scorers. They state that a certain pattern exists regarding strategy development. Inference and reliance on prior knowledge gradually declines (perhaps as learners' linguistic base increases), whereas the use of metacognitive strategies increases. The latter may be limited to more 'capable' learners and linked to the availability of processing capacity, which, in turn, may be related to linguistic knowledge. In their study, the participants are categorised as *lower-intermediate learners*. The reason is that they were at that time preparing for a lower-intermediate examination, the Advanced Subsidiary examination.

> Two students out of a group of 15 from three schools in England are the focus of the present study. All 15 were preparing for a lower-intermediate examination, the Advanced Subsidiary (AS) examination, at age 17 and had been studying French for 5–6 years. Immediately prior to the AS course, they had gained one of the top three grades in French for the examination taken at age 16, the General Certificate of Secondary Education (GCSE). On average, students received 4–5 hours of French instruction per week during the AS course, with formal listening occupying approximately a quarter of this time or less.
>
> (Graham, Santos and Vanderplank, 2008, p. 52)

Graham, Santos and Vanderplank (2011) also investigate the development of the listening proficiency and strategic behaviour of 15 lower-intermediate learners of French in England for six months with two methods, i.e. recall protocols and strategy elicitation. Firstly, the participants listened to two different audio recordings on the same topic and were asked to write in English everything they had understood. Then, they listened to four different texts and had to answer multiple-choice questions in English for strategy elicitation to capture participants' usual way of listening. They were requested to verbalise how they were about comprehending the text and answering the questions as fully as possible. The six-month study confirms that the use of metacognitive strategies increases with higher listening proficiency and that both inferencing and reliance on prior knowledge appear to become less prominent as learners' listening proficiency increases. In their study, the participants are categorised as again *lower-intermediate learners*. The reason is that they were at that time preparing for the AS examination.

> The students who form the focus of this article were drawn from a sample involved in a larger study (reported elsewhere, for example, in Macaro et al., 2006), including learners from four schools in England. All learners were preparing for the AS examination, taken at age 17, with 5–6 years of previous French study. In the larger study, 34 pupils completed a recall protocol listening task at Time 1, and 32 completed it at Time 2, approximately six months later (the lower figure arising from student absence or withdrawal from the course). Additionally, 23 students, chosen by convenience sampling, completed an individual listening activity for the purposes of strategy elicitation, again at Time 1 and Time 2. This study reports on the listening strategies employed by 15 students, selected because complete sets of data existed for them at both time points in the form of a recall protocol and an individual listening task.
>
> (Graham, Santos and Vanderplank, 2011, p. 439)

In both studies, the AS examination is referred to categorise the proficiency levels of the participants. However, the details *how* they were classified are not described.

It is a great hold back that standardised language proficiency test is rarely used in the field of EFL listening. Actually, no standardised language proficiency test was used in the following studies: Fujiwara (1990), Bacon (1992a, 1992b), Laviosa (1992), Goh (1997, 2000), Vandergrift (2003), Zhang and Goh (2006), Graham, Santos and Vanderplank (2008, 2011), Graham and Macaro (2008), Cross (2009, 2010) and Vandergrift and Tafaghodtari (2010). Rubin, Quinn and Enos (1988) employ the California Assessment Program in their study. However, it is a test of reading, writing and basic mathematical skills, but not a test to assess listening comprehension. Thomson and Rubin (1996) use the speaking ability section from the ACTFL. Vogely (1995), Ozeki (2000), Shirono (2003), Carrier (2003) and Suzuki (2009) use some tests, but again they are not standardised language proficiency tests. Thus, it is almost impossible to scientifically and objectively compare the participants' comprehension levels of those experimental research with other experimental research results. Chang (2008) uses the Test of English for International Communication® (the TOEIC®) to define the participants' proficiency levels in his study but did not mention the basis of the definitions how his participants were chosen and categorised.

To encourage the use of standardised language proficiency tests, they must be easily accessible outside the designated district, either low cost or free and available on the Internet. Moreover, the scores of standardised language proficiency tests must be convertible to those of international language proficiency tests such as the TOEFL®[3], the TOEIC® or IELTS[4] (Table 1).

Table 1: *Comparison of Some Standardised Test Scores/Grades*[5)]

TOEFL®	TOEFL®iBT	IELTS	Cambridge	TOEIC®	EIKEN
677	120	9.0	CPE	990	
650	115	7.5-8.5		890	
600	100	6.7-7.0	CAE	660-810	Grade 1
550	80				
500	61	5.5	FCE	590	Grade Pre-1
470	52	5.0		450-520	
		4.5			
450	45	3.5-4.0	PET	310-380	Grade Pre-2 to Grade 2
		2.5-3.0	KET	220	Grade 3
		1.0-2.0	Young Learners		Grade 4 to 5

If participants' comprehension levels *before* an experiment are not determined objectively by standardised language proficiency tests, then the results of the study cannot objectively be compared with other research results. Furthermore, even when employing a standardised language proficiency test, Rubin (1994) proposes that the division of groups or participants should be clearly described: Although DeFilippis (1980) used a standard instrument, the rationale for selecting the point where she divided the group is not clear' (Rubin, 1994, p. 212). O'Malley and Chamot (1990, p. 11) define successful listeners as 'those who report the greatest frequency, variety and sophistication of language learning strategies'. However, the range of successful learners varies depending on the instructor, and participants can be categorised differently across studies. Therefore, it is essential that every study should utilise an independent measure of success.

The second important reason for employing standardised language proficiency tests is that it would be difficult or sometimes impossible to compare the results with those of other studies without such standardised language proficiency tests and regardless of how many studies are conducted, EFL/ESL[6)] listening would not progress meaningfully. Rubin (1994) expresses that 'most of the research results are based on listening comprehension measures that have not been standardised, making it difficult to compare results' (p. 199) and 'most studies use either teacher judgment, course level or performance on a non-standard test' (p. 206). In addition, she states that studies which do not utilise standardised language proficiency tests cannot provide firm conclusions, and comparisons can be problematic for determining proficiency. Rubin's point is supported by Mendelsohn (1995) who stated that '… there is a need for *diagnostic tests*[7)] to assess learners' proficiency levels' (p. 137). Furthermore, Berne (1993, 2004) also addresses the importance of defining the categorisation of participants:

> The lack of a common, standardized measure of listening proficiency across these studies is problematic in that it may diminish the generalizability of the findings …. Thus, we cannot be sure that each of these studies is measuring the same thing when assessing listening proficiency. In addition, listening comprehension performance may vary according to the task used to assess it.
>
> (Berne, 1993)

> Therefore, in order to enhance the generalizability of their findings, researchers may want to consider adopting a common set of well-tested, objective criteria for assessing listening proficiency …
>
> (Berne, 2004, p. 523)

Those who conduct any experimental research in the field of EFL/ESL listening should realise that a standardised language proficiency test is essential for enhancing the credibility of a study's findings. With a standardised language proficiency test, participants' proficiency level in EFL/ESL listening is objectively possible to measure prior to any experimental research. It also enables to compare many different research results objectively. With the results compared objectively, it enhances generalisability of the findings in the field of EFL/ESL listening.

Notes:

1) Anderson (2010) claims that language learning involves certain steps (i.e., perception, parsing and utilisation) and proposes a cognitive framework of language comprehension.
2) The American Council on The Teaching of Foreign Language
3) The Test of English as a Foreign Language®
4) The lnternational English Testing System
5) The TOEFL® iBT test measures the ability to use and understand English at a university level, i.e., listening, reading, speaking and writing skills for academic tasks. The International English Language Testing System (IELTS) is designed to assess the language ability of candidates who want to study or work where English is the primary language of communication. The IELTS is accepted by thousands of organisations in more than 135 countries. Cambridge = Cambridge English Language Assessment, CPE = Certificate in Proficiency in English, CAE = Certificate in Advanced English, FCE = First Certificate in English, PET = Preliminary English Test, KET = Key English Test and EIKEN = a test in practical English proficiency, which is Japan's most widely recognised English language assessment.
6) English as a Foreign Language
7) The emphasis was made by Mendelsohn (1995).

References

Bacon, S. M. (1992a). The Relationship between Gender Comprehension, Processing Strategies, and Cognitive and Affective Response in Foreign Language Listening. *The Modern Language Journal, 76(2)*: 160–178.

Bacon, S. M. (1992b). Authentic listening in Spanish: How learners adjust their strategies to the difficulty of the input. *Hispania, 75,* 398–412.

Berne, J. E. (1993). The effects of text type, assessment task, and target language experience on foreign language learners' performance on listening comprehension tests. (Doctoral dissertation, University of Illinois, 1992). *Dissertation Abstracts International, 53,* 2354A.

Berne, J. E. (2004). Listening Comprehension Strategies: A Review of the Literature. *Foreign Language Annals, Vol. 37, No. 4,* 521–531.

Carrier, K. A. (2003). Improving High School English Language Learners' Second Language Listening Through Strategy Instruction. *Bilingual Research Journal, 27, 3,* 383–408.

Chang, A. C-S. (2008). Listening Strategies of L2 Learners With Varied Test Tasks. *TESL Canada Journal, Volume 25, Issue 2,* 1–22.

Cross, J. (2009). Effects of listening strategy instruction on news videotext comprehension. *Language Teaching Research, 13,* 151–176.

Cross, J. (2010). Metacognitive instruction for helping less-skilled listeners. *ELT Journal, 65(4),* 1–9.

DeFilippis, D. A. (1980). *A Study of the Listening Strategies Used by Skillful and Unskillful College French Students in Aural Comprehension Tasks.* Unpublished doctoral dissertation, University of Pittsburgh. U. S.

Fujiwara, B. K. (1990). Learner training in listening strategies. *JALT Journal, 12: 2,* 203–217.

Goh, C. (1997). Metacognitive awareness and second language listeners. *ELT Journal, 51(4),* 361–369.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems, *SYSTEM, 28,* 55–75.

Graham, S., & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of French. *Language Learning, 58,* 747–783.

Graham, S., Santos, D., & Vanderplank, L. (2008). Listening comprehension and strategy use: a longitudinal exploration. *SYSTEM, 36, 1,* 52–68.

Graham, S., Santos, D., & Vanderplank, L. (2011). Exploring the relationship between listening development and strategy use. *Language Teaching Research, 15(4),* 435–456.

Laviosa, F. (1992). *A preliminary investigation of the listening problem-solution process and the strategies of five advanced learners of Italian as a second language.* Unpublished doctoral dissertation. State University of New York, Buffalo.

Lowe, P. (1982). *The ILR Handbook on Oral interview Testing.* Washington, DC: Defense Language Institute.

Mendelsohn, D. J. (1995). Applying Learning Strategies in the Second/Foreign Language Listening Comprehension Lesson. In D. Mendelsohn, & J. Rubin (Eds.), *A Guide for the Teaching of Second Language Listening* (pp. 132–150). San Diego: Dominie Press.

O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics, 10,* 418–437.

O'Malley, J. M., & Chamot, A, U. (1990). *Learning Strategies in Second Language Acquisition.* Cambridge: Cambridge University Press.

Ozeki, N. (2000). *Listening strategy instruction for female EFL college students in Japan.* Tokyo:

Macmillan Language House.

Rubin, J. (1994). A Review of Second Language Listening Comprehension Research. *The Modern Language Journal Volume, 78, Issue 2,* 199-221.

Rubin, J., Quinn, J., & Enos, J. (1988). *Improving Foreign Language Listening Comprehension.* Washington, DC: U. S. Department of Education, International Research and Studies Program. (ERIC Document Reproduction Service No. 017AH70028)

Shirono, H. (2003). 言語学習ストラテジーを織り込んだ指導でリスニング力を伸ばすアクション・リサーチ [Action Research on Language Learning Strategies to Improve Listening Comprehension in English]. Japan: 三重県桑名高等学校セルハイ研究レポート [Mie Prefectural Kuwana Senior High School Super High Research Report].

Suzuki, K. (2009). リスニング・ストラテジー指導による EFL 学習効果 [Will listening strategy teaching help EFL learners ?]. *Dialogue, 8,* 20-37.

Thompson, I., & Rubin, J. (1996). Can strategy instruction improve listening comprehension ? *Foreign Language Annals, 29,* 331-342.

Ueda, M. (2015). *Towards Effective Teaching Methods in EFL Listening for Intermediate Learners,* 広島：渓水社.

Vandergrift, L. (2003). Orchestrating Strategy Use: Toward a Model of the Skilled Second Language Listener, *Language Learning, 53, 3,* 463-496.

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 leaners how to listen does make a difference: An empirical study. *Language Learning, 60,* 470-497.

Vogely, A. (1995). Perceived strategy use during performance on three authentic listening comprehension tasks. *Modern Language Journal, 79,* 41-56.

Wu, Y. (1998). What do tests of listening comprehension test ? — A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing, 15, 1,* 8621-44.

Zhang, D., & Goh, C. (2006). Strategy Knowledge and Perceived Strategy Use: Singaporean Students' Awareness of Listening and Speaking Strategies. *Language Awareness, Volume 15, 3,* 199-119.